attachment to ~~a~~ a letter to Mr. Travis Hoke
(ed: Popular Science Weekly) dated 21/1/1920

One of the most extraordinary of our cipher experiences, ~~except~~ from a psychological standpoint, was that in connection with some conspiracy cases against Hindus and Germans in late 1917. The men had been apprehended in this country although their crime was against our ally, Great Britain--that is, the crime of attempting to stir up a revolution in India. ~~in the late summer of 1917.~~ A representative from Scotland Yard came to us with an introduction from the Chicago Department of Justice. He brought with him a great stack of correspondence which the British postal censors had intercepted. Of information he could give us none except a list consisting of a few names of the suspects. This stack of papers, constituting letters passing between various agents, presented an appearance at first quite baffling. The letters consisted of series of figures in groups of three, such as 7-11-3; 8-5-6; 3-9-15; etc. After preliminary study it was determined that these figures referred respectively to page, line, and position in the line of a certain book. That is, 7-11-3 meant page 7, line 11, third letter in the line of this certain unknown book. The baffling part of the problem, of course, was what seemed to be the extremely remote possibility of reading any of these letters without having the book from which the numbers were taken. For a time it looked like an insurmountable task. However, after further study the fact was noted that the correspondents had repeated short sequences ~~series~~ of numbers in many cases. When desiring to encipher a word such as "committee", they would use as the indicators ~~given letter, such as "d" for example, a former~~ for "comm" the sequence of numbers taken from a word like "communication" series of numbers ~~referring to that letter would be used instead of citing~~ which occurred in the book, instead of using separate indicators taken ~~a different occurrence for the letters and a resulting difference in the~~ from different letters of different words scattered here and there throughout the book. ~~series of numbers.~~ After all it was something that might be expected with such a laborious system for concealment ~~of concealing a given message~~ that the correspondents would thus ~~make repetitions~~ save labor and search.

After considerable study we finally found a place where we thought it possible to fit the name of one of the Hindu agents. From this, fragment by fragment, the plain text of the messages was constructed and not only the plain text of the messages, but we could state with reasonable certainty that on page so and so of the unknown book used, line so and so of that page, appeared the word "Germany"; on another page and line, "government", and elsewhere, "constitution" and the like. From such evidence as that we made the deduction that the subject matter of this unknown book was political economy or history dealing with Germany & the German people.

As the time approached for the trial to be held in Chicago ~~of~~ certain of the agents, we were ~~[illegible]~~ informed by the Department of Justice that it would be extremely advisable to have on hand a copy of the book in order to convince the jury that the messages had not been read by witchcraft and black magic. We, therefore, instituted a search throughout the United States and in England among book dealers, asking for a book in English, quite modern in date, subject political economy or history of Germany, actual title and author unknown, but which would have on certain pages and in certain lines definite words which we listed. In every case the book dealers to whom the request was made replied in the negative. It seemed like a wild-goose chase.

The trial opened. The attorney supplied the courtroom with a large blackboard in order that the steps in the solution might be demonstrated to the jury in lieu of the lack of the book used. The second day of Mr. Friedman's wait ~~to be~~ called to the witness stand, as he was walking down Clark Street and past McClurg's, Chicago's largest book shop, on a moment~~'s~~ inspiration, he walked in and went to the section on political economy and rummaged among the books there. Imagine his astonishment at picking up the very book which the previous nation-wide search had failed to produce! An hour's labor devoted to checking up the solution we have reached without the book disclosed that we had done it correctly to the extent of over 95%!

786

A second incident, almost uncanny in its similarity psychiologically to the former, occurred in connection with the trial of more than one hundred Hindus and Germans. This trial was held in San Francisco in December of 1917. The correspondence of many of the indicted agents had presented an appearance somewhat similar to the correspondence in the former case; that is, the numbers were grouped in series of threes. In this case, however, the middle number of each group of three was always either number one or two, indicating a book divided into columns, two to the page. In addition, a preliminary study ~~indication~~ showed that the numbers referred not to individual ~~English~~ letters but to whole words, therefore, showing that the book used was a dictionary. Thus the nature of the book was determined as the correspondence was in the process of decipherment. ~~that is,~~ We felt that the book used was the second volume of a German-English dictionary and that it was many years old, certainly preceding 1900 in date. Again we telegraphed and cabled American and English book dealers for the English-German volume of a dictionary of such a date conforming to certain given requirements. Again the reply came back in every case in the negative.

Mr. Friedman departed for San Francisco. During the days after his arrival there, while waiting to be called to the witness stand, he continued his search for the book in question. He visited the university library ~~at Stanford and at~~ Berkeley, as well as all of the book shops in the city proper, to no avail. On a second ~~the~~ day on which he visited the University of California at Berkeley he met quite by accident on the campus a former instructor at Cornell who was then teaching at the California University. This man suggested that Mr. Friedman try the cooperative book store, the "college co-op". The two men then proceeded to this venerable book shop and inquired for the desired volume. Many dusty volumes were excavated in an effort to locate the book. The search was just resigned as hopeless when an elderly employee of the shop spoke up saying, "Wait a minute". He retired into the dim and remote corners from whence he issued with--believe it or not--an English-German dictionary of 1880 which answered in every detail the requirements, and was actually the volume so long sought for.

Once again the testimony was strengthened by having at hand at the time of the trial a copy of the book which had actually been used by the correspondents.

This last mentioned trial was a more or less sensational occurrence because of the fact that one of the Hindus, [redacted] who had turned state's evidence, was shot dead in the courtroom by another Hindu prisoner, [redacted] who in turn was shot dead by the United States marshal who fired from the ~~gallery~~ rear of the courtroom over the heads of the spectators.

The Hindus employed three different cryptographic methods. The first was what cryptographers call a pure cipher method using a keyword. It was as follows:

The numerical values of the letters of the message to be enciphered were obtained from the following diagram by taking the coordinates of the small square occupied by each letter. Thus, the message COME AT ONCE would be written as: 13 - 31 - 26 - 15 - 11 - 36 - 31 - 27 - 13 - 15.

	1	2	3	4	5	6	7
1	A	B	C	D	E	F	G
2	H	I	J	K	L	M	N
3	O	P	Q	R	S	T	U
4	V	W	X	Y	Z		

After the message was written out in this fashion, the numerical values of the letters of a keyword were added to the numbers. The sequence of key-numbers would be repeated as many times as necessary, to cover the entire text of the message. Suppose the prearranged keyword were the word LAMP ; its letters have these values (according to the diagram above):

L A M P
25 - 11 - 26 - 32

The values were then added to the values of the letters of the message. Thus:

C O M E A T O N C E
13-31-26-15-11-36-31-27-13-15
25-11-26-32-25-11-26-32-25-11
38-42-52-47-36-47-57-59-38-26

The message was then broken up into groups of five figures, ready for transmission. Thus:

38425 24736 47575 93826

The solution of messages enciphered according to this scheme, even though latter was wholly unknown to us, was not very difficult. The first step was, of course, to take the 5-digit groups and rewrite them into pairs of digits. Then the text was carefully examined for repetitions. The mathematical relations existing among the intervals between these repetitions gave clues to the length of the keyword. Once this length was determined the text was regrouped into "periods" or "cycles", the individual members of which obviously belonged to similar alphabets. For example, in the illustration above, the period is four, that is, there are four alphabets involved. Therefore, the 1st, 5th, 9th ... numbers belong to the first alphabet; the 2nd, 6th, 10th ... numbers belong to the second alphabet; the 3rd, 7th, 11th ... numbers belong to the third alphabet; the 4th, 8th 12th ... numbers to the fourth alphabet. After the numbers composing the text had thus been allocated to their respective alphabets, the latter were examined in detail, in order to solve the values of the various numbers belonging to each alphabet. These values were different, of course, in each of the four alphabets because of the key numbers that had been added. Thus, in the first alphabet the letter A was represented by the number 36 because this value was produced by adding the value of the first letter of the keyword L, (= 25) to the value of A (= 11). Of course, we did not know this when we were solving the message. We first solved the summation values (36 = A, for example), by applying the principles of frequency, and after we had read the whole message we were interested enough to try to find out why A had the value 36 in the first alphabet, the value 22 in the second alphabet, and so on. It was not very long before we had reconstructed the diagram shown above and had found out that the keyword used was the word LAMP.

The second method used by the Hindus was what is technically termed a code method. It made use of an ordinary dictionary of the two-column type. In order to indicate a word of the message, the number of the page, column and position in the column that the word occupied was set down. For example, the message COME AT ONCE might be indicated thus:

COME	AT	ONCE
78 - 2 - 48	35 - 1 - 17	122 - 1 - 52

Not knowing what dictionary had been employed, how did we proceed? Well, again by recourse to principles of frequency. Of course, the form of the groups, the fact that the central figure in each case was always either a 1 or a 2, immediately suggested that an ordinary dictionary had been used. Then all the groups were listed numerically, to find out what repetitions there were. The message was quite a lengthy one, running into several hundred groups. Now everybody knows that the letters composing written English are used with greatly varying frequencies, E being occurring used the most often, T, the next often, and so on. Letters also combine to form certain very common pairs of letters, such as ER, RE, and common endings, such as ING, ENT, and so on. But not so many people know that the same phenomenon is true as regards the frequencies with which words are used, as well as the combinations which they frequently form. Detailed studies have been made along these lines by various investigators, one of the latest of such studies being that made by Professor Godfrey Dewey of Harvard University. He found from an analysis of written text totalling 100,000 words that while there were 10,161 different words among the 100,000 examined,

10 words made up 26,677 of the whole 100,000
100 words made up 54,503 of the whole 100,000
1000 words made up 78,536 of the whole 100,000

The first ten words in his list were these (in their relative order of frequency):

THE, OF, AND, TO, A, IN, THAT, IT, IS, I.

You can't convey much intelligence using only these words, and yet you can't construct a long, intelligible, unambiguous message without using them over and over again. It is upon these words (and upon the word STOP, to the use of which most senders of telegrams seem so addicted) that the cryptanalyst first trains his guns when he is trying to shatter the wall behind which the intelligence conveyed by a code message is hiding. The code groups representing these words manifest themselves by their frequency of occurrence, by their positions in the message, and by their external form, when a dictionary is used. The last factor requires a bit of explanation, perhaps. In a dictionary the word AND comes far in advance of the word THE, for example. Consequently, the code number for AND will be much smaller than that for THE. The code number for the word YOU, on the other hand, will be higher than that for THE, and so on. In fact, the dictionary can be blocked off into 26 sections corresponding to the numbers of words beginning with each different letter of the alphabet, as follows:

Words beginning with A form 6.43 %

Words beginning with B form 5.35 %

Words beginning with C form 9.82 %, and so on.

Suppose the numbers in a code message of this sort extend from 1 to 100. Then it follows that a code number falling within the limits 1 to 6 will represent a word beginning with A; a code number falling with the limits 7 to 12 will represent a word beginning with B, and so on.

Again, suppose it has been definitely established that the code number 9~~7~~9-2-14 represents the word YOU; the code number 99-2-17, which may occur several times, must represent a word, certainly not far beyond YOU, and since there are but two or three relatively uncommon words (YOUNG, YOUTH, etc.) intervening between YOU and YOUR, the word is probably YOUR. This merely gives a general idea of the principles brought to bear after the "nonessential" words, such as THE, AND, STOP, YOU, AT, IN, ON, FOR, FROM, etc., have been found. The~~se~~ latter may not contribute much to the sense of the message, as nouns and verbs do, but for the cryptanalyst they are anything but "nonessential".

By following these principles we read about 95% of the whole of the long message given us for decoding. After the message was read, it became possible to try to determine exactly what dictionary was used, since we knew that if, for example 9~~7~~9-2-14 represented YOU, then we should find a dictionary in which this word was located on page 9~~7~~9, the right hand column, the 14th word in that column. That represented merely a painstaking search, interesting only for itself, certainly not essential for the general intelligibility and sense of the message but only for the corroboration of doubtful values. While waiting to be called to the stand during the trial of the Hindus in San Francisco in December of 1917, by a peculiar coincidence, I found the dictionary in the Cooperative Book Store of the University of California at Berkeley. It was the 2nd volume (English into German) of a rather old dictionary, such as students use in German courses. I have forgotten its name. Upon checking up our work I was pleased to note that we had made but one or two relatively insignificant errors.

The third method employed by the Hindus was perhaps the most interesting from the point of view of its solution. Here is how the rather long message began:

Dear
1-2-7
1-1-4
1-6-2
2-1-6
2-3-3
2-1-9
Am back from
7-12-26
7-1-19
7-2-6
7-1-19
7-1-12

Had lots
22-7-6
22-7-7
22-7-10
22-7-8
22-7-14
22-7-12
22-7-9

etc., etc.

This could be nothing but cipher, built up by taking the letters from a book and indicating what letters were taken by giving the number of the page, line, and letter in the line. For example, the very first set of numbers, 1-2-7, obviously meant that the letter taken was the 7th letter on the 2nd line on the 1st page of the a book. But what book?

Well, maybe we won't need the book at all. Look at the groups after the words HAD LOTS. The writer has obviously just returned from some place where he "had lots" of ________, what? Well, let us assume TROUBLE (these boys are always complaining about something or other).

If we assume the word is TROUBLE (note that there are just enough letters represented in cipher to fit the case), then

22-7-6 = T
22-7-7 = R
22-7-10 = O
22-7-8 = U
22-7-14 = B
22-7-12 = L
22-7-9 = E

And, if that is true, does it not follow that in the book on page 22, line 7, beginning with the 6th letter, there should be the following:

	6	7	8	9	10	11	12	13	14
Page 22, line 7:	T	R	U	E	O		L		B

The word TRUE stands out; we may speculate as to what follows it in the book! How about TRUE OF? If that is correct, then we have a new value, 22-7-11 equals F! Does it appear elsewhere in the message? It does, later on in the following word:

22-7-11
22-7-13
22-7-8
22-7-12
22-7-6

Suppose we substitute the values we now have.

22-7-11 = F
22-7-13 = ?
22-7-8 = U
22-7-12 = L
22-7-6 = T

What's the word? Why FAULT, of course, and that gives us one more value, 22-7-13 equals A, and enables us to add to our book!

	6	7	8	9	10	11	12	13	14
Page 22, line 7:	T	R	U	E	O	F	L	A	B

How about LABOR? That would give us two more values, 22-7-15 equals O, and

22-7-16 equals R. And so on, adding a word to the decipherment gives us a word to add to the book, the latter gives us a value or two to add in the decipherment, in an ever broadening stream.

Enough has been shown to give a clear idea of how we proceeded. Had this particular Hindu not been so lazy, had he not left what he thought were unimportant words in plain English, had he not taken letters close to one another in the book, the job would not have been so easy. In fact, had he been quite careful and skipped hither and yon throughout the whole book, never taking more than a single letter or two from a line, we would never have solved the message. But he was careless, and he was lazy, and he left us plenty of clues - enough so that we read practically all of the message without even knowing what book had been used. But we could go further. We could find the book, perhaps maybe? How? Well, haven't we good clues from the nature of the words and phrases we built up as we deciphered the message? For example, you saw how we established the fact that on page 22, line 7 of the book there appears the phrase TRUE OF LABOR. When we had finished the decipherment, we had built up about fifty or more words and phrases, such as these: CHARLEMAGNE, KAISER, FOREIGN OFFICE, GERMANY, RUSHED TO A NEWSPAPER OFFICE, HISTORICALLY, etc. Moreover, we could state the exact place in which each of these words and phrases appeared in the book, which from the nature of the text seemed to be one dealing with history or political economy, probably as concerned the Germans. Well, we sent out a query to many bookdealers in this country, asking if they would try to locate such a book, having, for example, on page 22, line 7, the phrase TRUE OF LABOR. It was, of course, too much to expect, and no answers were received. However, shortly before the trial opened in Chicago, when the Hindu who had written the message was confronted with our decipherment and saw that it was correct, he broke down and confessed. He even named the book he had used. It was Price Collier's "Germany and the Germans." A hurried visit to McClurg's, the big book shop, and I was soon corroborating our work. I found it to be correct to the extent of over 95%. Only in one word had we erred - and that was of less consequence than it was amusing. But I cannot tell you what the word was, because it would hardly pass the editorial blue pencil.